

## Migrating from SANergy to StorNext

As part of the installation and activation of the StorNext file system, data that currently reside on SANergy need to be migrated to StorNext. This includes the Data Pool science and browse collections, the OMS FTP pull area, the HEG download area, and any data the DAACs may have placed on SANergy.

Two alternatives have been proposed for performing the migration. This paper discusses the first of these alternatives.

### 1. Step-wise Migration by Collection

In this approach, both SANergy and StorNext are mounted such that ECS production can use both at the same time. ECS will operate as if the Data Pool was distributed over two file systems. Initially, all data is on one file system (SANergy). When the copy operation for a collection completes, operational use of the collection switches to its new location (on StorNext). Both copies will remain accessible to ECS, but users will be directed to the “current” copy to avoid confusion.

### 2. One Step Migration and Switch-over

In this approach, all data is copied from SANergy to StorNext while only SANergy is used by ECS production until all copy operations are completed. The system is then shut-down for a few hours and operation is switched over to StorNext. The transition script includes provisions that ensure that any ongoing inserts and deletions/cleanup are propagated to StorNext prior to switch-over.

In both approaches, data would remain on SANergy for some time after the copy is complete to allow for verifying the result of the copy operation, and to avoid the deletion overhead.

## Step-wise Migration by Collection

The following discusses issues and approaches for step-wise migration. Note that many issues are raised by both approaches, however, at different points in the transition. For step wise transition, most issues arise immediately, at the start of the transition. For one-step transition, most issues arise at the end of the transition period, prior to the final switch-over.

There are two pre-requisite patches that must be installed in OPS for the step-wise transition:

1. The Data Pool multiple file systems patch. This patch supports concurrent operation of several Data Pool file systems. It is scheduled for delivery in early August.
2. The Data Pool migration patch. This patch provides migration scripts and code patches required to support the migration.

## **Data Pool Directory Layout**

Data Pool file systems will be mounted below the mode-specific branch of the directory structure. As a result, a Data Pool file system is dedicated to a particular mode.

File systems for modes other than OPS are mounted outside the ftproot path and will be invisible and inaccessible to external anonymous users accessing the Data Pool via the Data Pool ftp service. However, the OPS file systems must be mounted below the Data Pool ftp root to permit anonymous ftp access to the data pool, and allow users to pull data via the OMS Synergy IV FTP Pull area.

For example, if the OPS data pool is spread over several file systems (called FS1, FS2 in the example), the directory structure will look something like this:

```

/datapool/OPS/user/FS1/<collgrp1>/<coll1>/...
                               /<coll2>/...
                               /<coll3>/...
                        /<collgrp2>/<coll4>/...
                               /<coll5>/...

/datapool/OPS/user/FS2/<collgrp0>/<coll6>/...
                               /<coll7>/...
                               /<coll8>/...
                        /<collgrp1>/<coll9>/...
                        /<collgrp2>/<coll10>/...
                               /<coll11>/...

```

etc.

All files for a given Data Pool collection must reside on the same file system. Different collections can reside on different file systems, even if they belong to the same collection group. The example above illustrates this. Collections from <collgrp1> and <collgrp2> reside on both FS1 and FS2, for example, <coll9> is a collection in <collgrp1> residing on FS2, whereas <coll1>, <coll2> and <coll3> from the same collection group reside on FS1.

To permit SANergy and StorNext to be used operationally by the Data Pool and OMS software, both SANergy and StorNext file systems must be mounted in the above fashion. This will be accomplished as follows.

The current SANergy Data Pool has the following directory structure:

```

/datapool/OPS/user/<collgrp>/<coll>/etc.

```

To make that structure fit the Data Pool directory rules, the SANergy path

```

/datapool/OPS/user

```

will be mounted as

```

/datapool/OPS/user/SAN

```

The /datapool/OPS/user directory represents the mount point for all the data pool file systems and must be on the root disk of every platform that has access to the Data Pool.

### ***Absolute Path References***

The above directory adds a new path element for the target file system into the absolute path to each file, as well as into the relative path from the ftproot. As a result, file pathnames for OPS change even before any collections are moved. For example:

/datapool/OPS/user/MOOT/MOD28L2.001/2004.06.17/sciencefile.hdf

/datapool/OPS/user/BRWS/Browse.001/2004.06.17/browsefile.jpg

become

/datapool/OPS/user/SAN/MOOT/MOD28L2.001/2004.06.17/sciencefile.hdf

/datapool/OPS/user/SAN/BRWS/Browse.001/2004.06.17/browsefile.jpg

As a result, absolute path references to the old path would become invalid. Such old path references are used, for example, when the Synergy IV OMS submits a physical media distribution request to the PDS. The request message includes the absolute paths for the location where the OMS staged the granules to the Data Pool.

To support such path references, the transition script will create symbolic links that provide shadow directory structure below the ftproot that mimics the existing Data Pool directory structure, but each directory entry will point to the real location of the collection.

For example:

/datapool/OPS/user/MOOT/MOD28L2.001 => ../../SAN/MOOT/MOD28L2.001

/datapool/OPS/user/BRWS/Browse.001 => ../../SAN/BRWS/Browse.001

The links will point to the current location of the collection on SANergy. After the collection was copied, the link will be updated to point to the new location of the collection. For example, after MOD28L2.001 was copied to FS1 on StorNext, the link would be updated to:

/datapool/OPS/user/MOOT/MOD28L2.001 => ../../FS2/MOOT/MOD28L2.001

As a result, absolute file references that were constructed before the transition started will remain valid, and will continue to be valid even after the collection was copied.

### ***Relative Path References***

Many of the existing path references are relative to some location below the ftproot.. Examples are:

browse links:

/datapool/OPS/user/SAN/MOOT/MOD28L2.001/2004.06.17/browselink.jpg

=> ../../BRWS/Browse.001/2004.06.17/browsefile.jpg

FTP Pull links<sup>1</sup>:

```
/datapool/OPS/user/SAN/pulldir/orderdir/pulllink
=> ../../MOOT/MOD28L2.001/2004.06.17/sciencefile.hdf
```

HEG links<sup>2</sup>:

The existing relative path references remain valid as long as the source and target location are both in the SANergy file system. If the source location is changed to a different file system, the link must be updated. For example, when the MOD28L2.001 collection is copied to StorNext, any links to browse files that may exist in the collection subdirectories must be re-created at the new location:

```
/datapool/OPS/user/FS1/MOOT/MOD28L2.001/2004.06.17/browselink.jpg
=> ../../SAN/BRWS/Browse.001/2004.06.17/browsefile.jpg
```

On the other hand, FTP Pull links do not have to be updated until the FTP Pull area itself is copied to StorNext or the collection is deleted from SANergy. This is because:

- Any FTP Pull links that are created after a collection was moved to StorNext will point to its StorNext location (see also Footnote 1).
- Any FTP Pull links created before that time will point to the original files on SANergy and will remain valid until those files are deleted.

As a result, the transition script will update the FTP Pull links only when the FTP Pull area is moved.

Orders for HEG processing via the Data Pool involve two types of file references. One is in the table `DlCartItem (inputfile)`. It provides the path to the input file referenced by a particular line item in the HEG order. The format of these links is:

- **TBD#01**

The other type of link is in the HEG download area and points to the HEG output for each line item. The format of these links is

- **TBD#02**

---

<sup>1</sup> The FTP Pull link reflects the OMS behavior as of this writing. However, a NCR has been filed to make the location of the OMS FTP Pull area configurable, and to include the file path for a collection (when configured) in the link:

```
/datapool/OPS/user/SAN/pulldir/orderdir/pulllink
=> ../../SAN/MOOT/MOD28L2.001/2004.06.17/sciencefile.hdf
```

The NCR must be implemented before collections can be copied to StorNext, to ensure that the FTP Pull links created after the collection was copied to StorNext will point to the new location..

<sup>2</sup> A NCR has been filed on the HEG links. To support multiple file systems and StorNext, HEG links must be converted to soft links, and must use the file system path to ensure that links that are created after a collection was copied to StorNext will point to the new collection location.

Neither link type of will need to be updated until the HEG download directory is moved. This is because:

- **(TO BE VERIFIED)** The HEG input file references that were created prior to the transition will remain valid because they do not contain file system references, and thus will be resolved correctly via the shadow directory tree which the transition script creates.
- **(TO BE VERIFIED)** HEG input file references created after the transition will include in the path the file system on which the collection is located. References created before a collection is moved will reference the old location. These files will be retained, or if they were removed later, the collection directory would be replaced by a link pointing to the new collection location. As a result, these references will remain valid.
- **(TO BE VERIFIED)** HEG input file references created after a collection has moved will point to the new collection location, and hence, will remain valid from then onwards.
- **(TO BE VERIFIED)** The HEG download directory as well as the output files which the links in the download directory reference will remain on SANergy until the end of the transition. As a result, these links will remain valid. As one of the last steps of the transition, the download directory and all related files will be moved to StorNext and the links will be re-created. This step will be performed while HEG ordering and processing are disabled.

### ***Retention of SANergy Collections***

The migration concept assumes that collections that were copied to StorNext will remain on SANergy until SANergy is decommissioned. This will allow time for checksum verification of the copied science files, ensuring that the science files can be compared with its source and can be re-copied if the verification fails. It also ensures that any accidental access that uses outdated path references will succeed as long as SANergy remains operational at that DAAC.

However, if a collection is deleted from SANergy prior to that, the removal process must create a link that points to the new collection location on StorNext after the collection directory was removed, to extend the validity of old references.

### ***Operations Concept***

Prior to data migration, the transition will create mount directories on the root disk of every Data Pool platform.

Next, the transition will create the shadow directory tree mentioned on the Data Pool platform (any other platforms are **TBD#03**). If the DAAC defines new collections during the transition time period, it is responsible for creating the corresponding link in the shadow directory tree. DAACs may wish to update group and collection links they established on this occasion.

Before SANergy can be mounted as a separate Data Pool file system, the existing Data Pool collections may need to be updated to indicate the file system on which they reside

(**TBD#04**). The file system entry for SANergy will be updated to reflect the SANergy mount, or added if it does not yet exist.

After SANergy was mounted in its new location, the transition will move the most recent insert listings from the current ftproot (on SANergy) to the ftproot directory on the wu-ftp host platform.

Data will be migrated from SANergy to StorNext using a migration script, in the following sequence:

1. Data Pool science collections
2. Data Pool browse collection
3. OMS FTP Pull area
4. HEG download directory
5. Other Data

Concurrently, a background checksum verification utility will verify a subset of the Data Pool science files that have been copied to StorNext. The subset will be configurable and will be large enough to ensure with high probability that all files were copied correctly.

The migration script will be able to operate on more than one platform, and migrate several collections concurrently. This will distribute the copy workload and improve copy throughput (**TBD#05**).

The DAAC will be able to

1. shut down and start-up migration activity to control when migration occurs (migration scripts will be re-startable)
2. control how many copies of the migration script will be executing and on which platforms, and change this as operational considerations dictate
3. limit the current segment of migration activity to a specific set of collections, e.g., to limit the migration to some “test” collections during the initial migration phase
4. control the percentage of files that undergo checksum verification
5. startup and shutdown verification and control the number of concurrent verification processes

The operational procedures for these steps will be documented and provided to the DAACs.

Migration of science collections will occur in parallel with normal operation, subject to throughput constraints imposed by concurrent operational workload.

~~Whether the same applies to the migration of Browse directories is TBD#07.~~ Migration of Browse, FTP Pull and HEG directories will occur during downtime.

The migration concept for other data is **TBD#08**.

It is conceivable that Data Pool ftp and web users encounter temporary problems when trying to access a moving collection (see below for explanation). The DAACs should

include appropriate notices in the ftp welcome message and the Data Pool home page during the migration period.

After the migration is complete, SANergy will remain active until all verification activities have been completed.

**TBD#09:** As SANergy is de-installed, a temporary directory tree might be created that mimics the SANergy collection directory tree, but points to the current location of the collection, to extend the validity of old file/path references.

### ***Switching Back To SANergy Only Mode***

The data migration will be tested both in Landover and at the DAAC. Moreover, Stepwise migration reduces the risks as the DAAC would uncover any serious problems in the data migration or in the concurrent operation of SANergy and StorNext during the transition of the first few collections. [A switch back mechanism will be provided which will allow a DAAC to return to SANergy only operation \(TBD#11\).](#)

~~Still, whether to provide a switch back mechanism needs to be considered (decision is TBD#11).~~

In order to switch back to SANergy only mode, the initial setup steps ~~must~~ will be reversed by:

1. Moving the most recent inserts listings back to their SANergy location
2. Removing the shadow directory tree
3. Updating the file system pointers in the Data Pool inventory back to their original values
4. Remounting SANergy in its original path
5. Reverting to the original permissions on the SANergy directories for migrated collections so they can be browsed via FTP.

In addition, if any collections were migrated already, their recent inserts ~~must~~ will be re-applied to SANergy (since they now would be located on StorNext). Based on the earlier assumptions, the amount of data that must be reconstituted on SANergy will be small, and the switch-back should occur during down-time. There are two options for doing the switch-back:

1. Copying the newly inserted files from StorNext [back](#) to SANergy, and regenerating any newly created links. [This option will be supported by the data migration script.](#)
2. Removing the new inserts from the Data Pool inventory, and then re-inserting them from ECS. [This option is not supported by the migration script. Though possible, it would require manual steps by DAAC operations.](#)

In addition, any newly created links to the newly inserted files must be updated to point to SANergy rather than StorNext.

## ***Migrating Science Collections***

### **Migration Steps**

The following is a brief description of the steps which the migration script will perform for each science collection (**TBD#12**):

1. The script obtains the target StorNext file system for the collection.
2. The script updates a flag in the Data Pool inventory indicating that the collection is “moving”.
3. The script creates the collection directory on StorNext, setting the permission so its contents will be invisible to anonymous ftp.
4. The script copies the science files to their new StorNext location, and re-creates the browse links at the new location, as well. It does so in order of insert time (**TBD#13**).
- ~~5. The script checks the Data Pool for inserts of new files/links that it may have missed, and copies or re-creates them as well. It does this incrementally and in sequence of insert time.~~
- ~~6.~~5. The script checks the active Data Pool inserts to determine whether there are any insert processes in progress for this collection, and if so, may terminates them re-triably (see **TBD#14**).
- ~~7.~~6. The script updates the collection entry in the Data Pool to point to the new file system, and updates the permissions on the SANergy copy of the collection directory to make its contents invisible to anonymous ftp users.
- ~~8.~~7. The script makes one final check for missed inserts.
- ~~9.~~8. The script resets the flag in the Data Pool to indicate that the collection is no longer “moving”.
9. The script updates the shadow directory link for the collection to point to StorNext.
10. The script hides the contents of the SANergy collection directory from anonymous ftp, and unhides the contents of the StorNext collection directory.
11. The scripts executes some reconciliation steps to allow subsequent recovery of any failed copies

These steps are designed to minimize operational impact if a collection is accessed while it is being moved (see discussion below).

### **Recovery**

To complete the migration of a science collection that was interrupted prior to step 6, the operator needs to ~~restart the script-(TBD#10)~~. ~~Recovery if the interrupted migration will then be automatic: This will resume~~ migration of the collection is resumed, recognizing which files have already been copied successfully and will not copy them again. Note



that during this time, new Data pool inserts will be directed to the SANergy copy of the collection, because the StorNext copy of the collection is not yet Operational.

~~Postponing recovery is not recommended. However, it is possible should the need arise, but the DAAC will need to contact Landover support for instructions. To postpone the recovery, the operator will need to (TBD#16).~~

After completion of step 6, the Data Pool will recognize the new location of the collection, and new inserts will be routed to it. ~~Though it is possible to return operation to using the SANergy version of the collection, this is not recommended and will require assistance from the Landover support staff. If the operator wishes to revert to using the collection on SANergy, the operator must take the following TBD#17 steps.~~

## Operational Impact

This section describes the impact of migrating science collections concurrently with operation of the various ECS components that access the Data Pool.

### 1. Data Pool Insert

The Data Pool insert utility obtains the location of the file system into which to insert a granule from the Data Pool collection entry. It will point to SANergy before step 6 in the transition script, and to StorNext thereafter. The transition script will catch up with any inserts that occurred during the copy phase during its step 5.

To guard against concurrent inserts creating potential race conditions, the transition script will identify any current Data Pool inserts during Step 6 and cause them to fail retriably. This will cause the insert to be repeated (unless they happened to complete). The retry will insert the files into the StorNext file system. (whether this is necessary depends on a potential DPIU code change<sup>3</sup>).

After doing so, the transition script will check once more for any inserts that may have occurred into the SANergy version of the collection and copy them to StorNext, as well. This guards against the possibility that an insert completes just as the script checks the list of active inserts. Normally, this check will come up empty.

### 2. Data Pool Drill Down

The Data Pool web interface will disallow access to a collection while it is flagged moving. However, because the files are not removed from SANergy, old URL will continue to work even after the copy completes.

### 3. Data Pool OMS Orders

The Data Pool web interface will disallow access to a collection while it is flagged moving. However, pre-existing web orders will remain valid since granules are referenced by their Data Pool identifier, not by their pathnames.

### 4. Data Pool HEG Orders

---

<sup>3</sup> A NCR will be filed to have the DPIU check the file system path on a configurable basis just prior to the final data move into the Data Pool. That data move takes just a few seconds. The NCR fix will simplify the approach.

The Data Pool HEG order interface includes the path names of the input granules in its order. Since the existing granules will remain in the SANergy file system, those references will remain valid until the SANergy file system is de-installed. At that time, any remaining references to SANergy files must be updated to reference the new file location on StorNext.

Accesses by HEG are not affected by concurrent copying of the same files to StorNext. The HEG download directory is not affected by the migration of science collections. Migration of the input links and the download directory is discussed in a later section.

## 5. Data Pool Cleanup

The impact of running Data Pool Cleanup while migrating collections to StorNext depends on the relative timings of migration and cleanup steps. Both scripts query the Data Pool database to populate the list of files to be deleted / migrated. Cleanup then removes the granule inventory entries from the Data Pool, and then deletes the files from disk.

- If Migration obtains the list of the files to be migrated for a collection after Cleanup removed the cleaned-up granules from the inventory, they will not be included in the migration and when Cleanup completes, the SANergy and .StorNext copies will be consistent.
- If Migration obtains the list of files to be migrated for a collection, any of the files that Cleanup will later remove will be included in the migration. If they are deleted before Migration copies them, Migration will treat the like orphans, i.e., log the occurrence and continue. Otherwise, Migration will copy these files, and after both complete, the StorNext file system will include files that are no longer referenced in the inventory, i.e., ~~can result in orphans.~~
- If Cleanup obtains the list of files after the collection completed migration, it will perform cleanup on the StorNext file system, i.e., the StorNext file system will reflect the cleanup correctly. ~~The reasons is as follows: Cleanup will use the SANergy or StorNext version of the collection for cleanup, depending on when exactly the list of files that need to be cleaned up is created. If the file list references the SANergy version of the collection, files could be deleted while they are being copied, or after they were copied. In either case, the a copy of the file would appear in the StorNext version of the collection, but the granule entries would be deleted from the Data Pool inventory database, resulting in orphaned files.~~

The impact of running Data Pool validation while migrating a collection to StorNext is as follows:

- Orphan checking verifies that files older than a specified time period are still referenced in the Data Pool inventory database. No matter whether the file is found on StorNext or SANergy, it is an orphan. ~~However,~~ Removal of an orphan from SANergy ~~has no impact on StorNext.~~ The migration uses the Data Pool inventory database to construct the authoritative list of files to be copied, i.e., after migration, the StorNext copy of the collection will not have these orphans. ~~after it was copied to~~

~~StorNext will cause the orphan to remain in StorNext (and found on the next orphan check).~~

- Phantom checking verifies that there are no Data Pool inventory entries referencing files that are no longer on disk. If files are missing from disk for a collection that is migrating, they would be missing from the SANergy and StorNext copies; and phantom removal would drop them from the Data Pool inventory. The migration will log phantoms but ignore them otherwise (i.e., it will not remove them). However, phantom removal may create StorNext orphans if it removes files from SANergy after they were copied.

Orphaned files do not cause problems other than taking up extra space in the Data Pool. Orphaned files can be removed via a later Data Pool validation run. Data Pool validation can be constrained to orphan checking and to particular collection groups, but not to individual collections.

The time required for such a check, given the GDAAC Data Pool size is **TBD#18**.

As an alternative, whether it is possible to discover and handle such orphans while the collection is migrated is **TBD#19**.

#### 6. Data Pool Most Recent Insets Utility

The listings created by the utility do not include file system information. The listings for a collection will be copied by the transition script as it migrates the collection. The most recent insert listings at the ftproot level will be migrated at the start of the transition, and will be maintained at that location thereafter.

#### 7. Data Pool Log Analyzers

**TBD#25**

#### 8. Data Pool FTP Access

Concurrent copying of a file has no impact on its ftp pull. That is, users that currently download a file would not be aware of the fact that file is currently being copied from SANergy to StorNext.

Because of the shadow directories, users that use the current Data Pool path structure would see no change, regardless of where the file is located. If they are currently switched to a subdirectory of a collection that is being migrated, they would be able to continue downloads from that (SANergy) directory even after the migration of the collection completes. Depending on wu-ftp implementation, they may find it impossible to list the contents of such directories once the copy completes (**TBD#20**); however, if they re-navigate to the directory, they would now look at the StorNext version of the collection and would experience no such difficulties.

Users should be discouraged from navigating to collections via the file system directories. Any time a collection is migrated, this navigation path will change.

We recommend that DAACs publish the data migration schedule and possible impacts, so users, especially those using access scripts, are aware of what types of problems to expect and when they might occur.

## 9. OMS Staging

OMS performs staging via the Data Pool insert service. We do not foresee any impact on staging, except that any inserts that are failed retriably at the end of a collection migration would experience a delay.

OMS caches the file system / collection map in memory. The collection move utility leaves a link to the new location behind, which can will direct OMS to the correct file system until the cache was refreshed. The SAN migration utility cannot do this because it preserves the existing files until verification is complete. As a result, OMS must update its cache after a collection was copied, because new files will be inserted into the StorNext file system rather than SANergy<sup>4</sup>.

## 10. Synergy IV FTP Pull Orders

After all granules for a Synergy IV FTP Pull order have been staged, the OMS creates a directory for the order in its ftp pull directory, and populates it with links that point to the ordered granule files. Migration of the links has been discussed earlier.

FTP Pull download will not be impacted by the migration of science collections. FTP Pull links remain in the same place during that transition phase; and regardless of whether a link references the SANergy or StorNext version of a file, the user will be able to pull it.

The OMS queues actions to remove FTP Pull links when the order expires, and the related granules if they are not needed for other purposes. If the OMS performs FTP Pull cleanup for granules that belong to a collection that is currently moving, the granule may be removed from SANergy after it was copied to StorNext, i.e., would become an orphan on StorNext.

Whether this issue can be addressed by OMS rather than orphan checking is **TBD#22**.

## 11. Synergy IV FTP Push Orders

OMS constructs the path name for files that it needs to push at the start of the FTP push operation. There is a remote possibility that OMS might construct the pathname using the new file system location for the collection, but the file itself was not yet copied (during step 7 of the transition script). If this occurs, the ftp push operation will fail and the request will be suspended. Resubmitting the request will clear the problem<sup>5</sup>. To reduce the chance of such a problem, the transition script steps have been designed to reduce the amount of time needed for Step 7, and the chance that it will actually locate a file that has not yet been copied.

There is also a possibility that the OMS constructs the SANergy pathname for a file just as the migration completes. Since the original files on SANergy remain accessible to FTP Push, this will not impact the FTP push operation.

---

<sup>4</sup> A corresponding NCR has been filed.

<sup>5</sup> This is the subject of a NCR: OMS currently caches the file path for later use. Normally, this is efficient because it saves several database access. However, if the path cannot be found, OMS needs to look up each path element in the database in case the path has changed.

Concurrent copy operations will have no impact on the ftp push operations (other than throughput).

## 12. Synergy IV Physical Media Orders

OMS constructs the path name for files when it submits the order to the PDSIS. Orders that were submitted to the PDS before start of the transition will use pathnames that point to the shadow directory. They will be resolved to the current location of the collection. Orders that are submitted to the PDS after start of the transition may point to files on SANergy, but when the order is finally processed by the PDS, the collection may have been migrated to StorNext. As long as the files on SANergy are retained, there will be no impact on the order.

After the referenced files have been deleted (e.g., because SANergy was decommissioned), such an order would fail because the production module cannot find the input file. Failing the order, and then resubmitting it via the OMS will resolve the issue. This type of problem could occur after transition to StorNext is complete if physical media orders have been queued in the PDS for some time and reference collections that were migrated last.

It is recommended to set the OMS high watermarks for physical media orders such that no more orders are queued in PDS than can be processed within a few days. This ensures that after that time, no PDS order will have outdated path references.

## 13. Export of Data Pool URL to ECHO

Potential impact depends on the format of the URL exported to ECHO and is **TBD#21**.

## 14. Export of Data Pool URL to EDG

URL are returned to users as a result of searches. It is possible that a user receives a URL that points to the collection location on SANergy, but when the user employs the URL, the collection has migrated to StorNext. The old URL will be valid as long as the old data is retained on SANergy.

Once SANergy is de-installed, the URL will become invalid. That situation is very unlikely to occur if the URL is used within a few days. It is possible to prevent this situation by creating a second (temporary) shadow directory tree after SANergy de-installation that mimics the SANergy directory tree, but points to the correct collection location.

## ***Migrating The Browse Collection***

The browse files will be migrated during ~~(downtime/~~[system operation](#)~~)~~ **TBD#7a**. The browse file migration step will also update the existing browse links to point to the new location of the browse collection.

The number of browse files and links at GDAAC is ~~TBD#7b~~ [about 8,000 granules, respectively 16,500 links](#).

The throughput for Browse migration is **TBD#7c**. The estimated time for migrating the browse collection at GDAAC based on these numbers is **TBD#7c**.

### ***Migrating The OMS FTP Pull Area***

Migration of the FTP Pull area **will occur during system down time**. It does not involve the copying of files, but rather, the creation of order directories and links on the StorNext file system that will house the ftp pull area.

The number FTP Pull links at GDAAC is **TBD#23a** (estimate: <30,000; source: GDAAC Statistics compiled by Jean Bedet until Feb 2004).

The speed with which they can be created on StorNext is **TBD#23b**.

The estimated time for FTP Pull migration at GDAAC based on this numbers is **TBD#23c**.

~~While the new ftp pull area is created, the old ftp pull directories and links on SANergy could be used. However, OMS may remove expired ftp pull directories if it is operating during the ftp pull area migration. If this occurs during or after the copy of the ftp pull directory, the directory being removed is the SANergy copy; its StorNext counterpart would not be removed and there is currently no mechanism for automatic removal. Countermeasures are only relevant if transition during downtime is not possible.~~

### ***Migrating The HEG Links and Download Directory***

The DAAC may elect to perform the HEG transition during data pool downtime. At a minimum, however, HEG processing must have stopped and the HFE server must be shut down; HEG ordering must be disabled during the transition of the input links; and Users should be alerted that access to the HEG download area may be impaired while the migration is in progress.

The transition script will update the input file references in the HEG order information stored in the Data Pool database. From then on, these links will no longer rely on the shadow directory tree.

Next, the script will copy the download directories and their contents from SANergy to their new location on StorNext, and update the link in the order directories. The script then makes the old download directory unwriteable. The download directory on SANergy may need to be retained until SANergy is decommissioned and its directories are replaced by a SANergy shadow directory tree. (**TBD#24**).